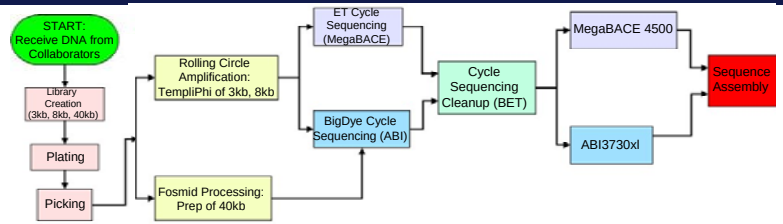
An Overview of the Joint Genome Institute Production Sequencing Line

Chris Wong, Miranda Harmon-Smith, Kathleen Lail, Chris Daum, and Lena Philip

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

ABSTRACT

The Department of Energy's (DOE) Joint Genome Institute (JGI) Production Genomics Facility (PGF) performs high-throughput Sanger sequencing using whole genome shotgun sequencing methods. The goal of the Production Line is to produce high quality sequence data in order to allow the downstream assembly of whole genome, BAC and metagenomic projects. The Production Line generates end sequence from three different size DNA fragments: 3, 8, and 40 kb, to a depth of 8.5x coverage. The production line is comprised of three subgroups: Library Support, Sequencing Prep and Capillary Electrophoresis, which consist of about 35 technicians who perform their work within roughly 12,000 square feet of laboratory space. The production line uses 71 Applied Biosystem 3730xl DNA sequencers and 36 MegaBACE 4500 DNA sequencers to generate approximately 90 Megabases a day. The Applied Biosystem 3730xl sequencers are operated 24 hours a day, 365 days a year, while the MegaBACE 4500 sequencers are run approximately five times a day, five days a week. For fiscal year 2006, the Production line produced 49 million lanes and 33 gigabases of sequence with an average read length of 647 bases per lane, sequencing approximately 200 projects. This poster will describe the infrastructure, process steps and quality control methods to ensure the production of high quality sequence.

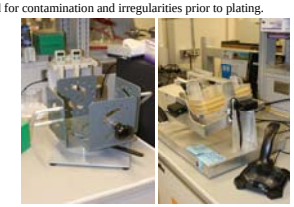


LIBRARY SUPPORT PROCESS

PLATING

Library Support receives transformation stocks from the Cloning Technology Group. The stocks are stored at -80°C until plated. Technicians are assigned multiple libraries and up to 40 bioassays on a daily basis. On average, 120-160 bioassays are plated per day. Using the Production Data Base, the technician will determine which Teknova antibiotic bioassays to use with each library they are plating: Carbenicillin w/ X-gal (for pUC, 3kb), Chlorampenicol 20 w/ X-gal (for pMCL200, 8kb), Chloramphenicol 12.5 w/ X-gal (pcc1fos, 40kb), or Kanamycin 30 w/ X-gal (various vector types and sizes). Each bioassay will create approximately two and one-half 384 well plates. The bioassays are QC'd for contamination and irregularities prior to plating.

Transformation stock stored at -80°C is thawed on ice and mixed with S.O.C. media based on the recommended plating volume for the given library.



Plating is done using a manual or fully automated plating device using 5mm glass beads to spread.

Bioassays are incubated for 18 hours at 37 °C.

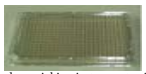
PICKING

Picking of individual colonies from the bioassays is an automated process using the Genetix Qpix and Qpix2 XT instruments. Our production line has 5 colony picking instruments that run for 16 hours per day. They produce on average a total of 300 destination plates per day. We sterilize the environment with two 30min. UV irradiation events, at the beginning and at the end of each day to eliminate possible contamination. We also perform routine sterility tests twice a day to ensure proper washing of the picking pins between each inoculation event and to eliminate the possibility of cross contamination from well to well.

Bioassays are pulled from incubators after 18-20 hours incubation and loaded onto a Qpix instrument.

384 well destination plates are filled with LB/glycerol + antibiotic using a plate crane and microfill instrument.

Colonies are imaged by the attached LCD camera and selected based on preset criteria.



The colony picking instruments pick and inoculate the colonies into 384 well destination plates. These plates are then labeled with barcodes specific to the library that was picked and tracked in the database.

The destination plates created are then incubated for 18 hours at 37 °C

OPTICAL DENSITY QC CHECK POINT

After picking and incubation, each 384 well plate is QC'd to ensure that adequate growth is obtained in every well. We use a SpectraMaxPlus spectrophotometer instrument to obtain the optical densities of each well. Ranges and colors are set to indicate the growth pattern in the wells for easy detection. Plates are stored at 4°C for next day processing or at -80°C for longer term storage.

Automated platecrane and dual Spectramax Plus Unit

Example of a 384 well map showing no growth and minimal growth wells

No growth → <0.071 (red wells)

Minimal growth →0.071-0.1 (yellow wells)

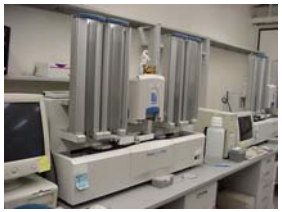
Good growth → 0.11-0.70 (green wells)

Overgrowth → >0.70 (blue wells) * These plates are not sent for sequencing because of possible contamination

SEQUENCING PREP PROCESS

Rolling Circle Amplification

The purpose of this step is to amplify circular DNA template for cycle sequencing.



Plates filled with buffer, along with the 384-well glycerol stock plates (containing *E. coli* with the DNA inserts) created in the Library Support Process, are loaded onto the Matrix PlateMate Plus robots. Glycerol stock contents are transferred to the buffer plate. The 384-well plates are placed onto thermocyclers, where cells are lysed and the plasmids are released. Next, the TempliPhi Premix containing: Phi29 Polymerase, Random hexamers and dNTP's are added. The plates are incubated overnight at 30°C, where many copies of the DNA is made.

Sequencing Chemistry (Sanger Reaction)

The Sanger reaction step is a modified PCR reaction that produces an enormous number of dye labeled DNA fragments.

The CyBio Well Vario combines the aliquot step and the dispense step. A chemistry cocktail, comprised of specific primers, Taq Polymerase, dNTP's, ddNTP's, buffer, & water, is added to each plate, after the RCA reaction. A forward and reverse plate is created, from each RCA template plate, and then run through 30 cycles on the thermocyclers, to complete the Sanger reaction.

Post Sequencing Clean Up

Before the plates can be loaded onto the capillary sequencers, the leftover reagents, cell debris, buffers, and salts must be removed from the sample. This process uses a modified magnetic bead protocol to purify DNA fragments from the sequencing reaction. This step is performed entirely on the Beckman Coulter Biomek FX robot.

The Biomek places the source plate onto the deck and adds the BET solution (200 proof ethanol, water, Tetra ethylenglycol, and washed Seradyne Magnetic Carboxylate-Modified Beads) to the plate. The DNA is induced to attach to the beads but not the excess terminators. The source plates are returned to the stackers and incubated. The source plates are then placed onto the magnets where the DNA/beads are drawn down so waste solution can be aspirated. A 70% EtOH wash step occurs, followed by an aspiration.

The plates are allowed to dry. Water is dispensed into the plate. The plates are then transferred to the magnets. Since DNA has a higher affinity for water, the beads are drawn down while the DNA remains in solution. Lastly, water + DNA is transferred to the destination plates and returned to the stackers. The plates are now ready to load on the capillary sequencers.

CAPILLARY SEQUENCING PROCESS

The JGI has chosen to utilize two different DNA analysis platforms. Each contributes to the overall sequencing effort of the department in a different manner. The two platforms utilized by the JGI are the MegaBace 4500-36 instruments and the ABI 3730xl-71 instruments.

The ABI3730xl was released in 2002. This highly automated platform allows for 24 hour a day, seven day per week processing of sample plates, with limited interaction by a small staff of technicians. The technicians load/unload samples daily and change out reagents a few times per week.

The 384 capillary array system of the MB4500 allows for high sample throughput on each sequencing run performed; however, it requires a technician to manually load each run. The MB4500 uses a high power solid-state laser, which is mated to a scanning confocal optics system, providing enhanced detection sensitivity for long reads; attaining 100 more bases per lane than the ABI3730xl platform (on average based on JGI run parameters). Additionally, the MB capillary arrays and solid-state laser achieve much longer operational lifetimes when compared to the capillary arrays and the argon-ion laser of the ABI3730xl platform. The ABI3730s have an operational uptime of 98.8% while the MB4500s have an operational uptime of 97.9%; downtime on both platforms is primarily due to instrument errors.

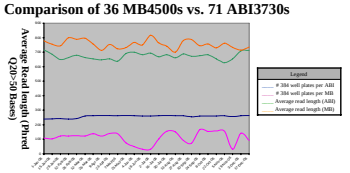
ABI 3730xl DNA Analyzer (71 sequencers)

MegaBace 4500 DNA Analyzer (36 sequencers)

Operational Parameters and Specifications for the MB4500 and ABI3730xl DNA Sequencing Platforms

Operational Parameter	ABI 3730xl Specification	MegaBACE 4500 Specification
Run Parameters	8.5kV for 70min	Variable
Injection Parameters	1.5kV for 35sec	2.0kV for 35sec
Operating Temperature	50°C	55°C
Sieving Matrix	POP-7 polymer	V2 LPA
Capillary Array Type	- 50 cm – 96 capillaries - 150 micron o.d. & 50 micron i.d.	- 70 cm – 384 capillaries - 200 micron o.d. & 75 micron i.d.
Maximum Sample Lane Throughput (based upon JGI Run parameters)	1445 samples/day/instrument	2304 samples/day/instrument
Laser power/type	25mW argon-ion	100mW solid state
Laser accessories (cooling)	External HVAC system with negative pressure hookup to draw air from system	None - no external Power Supply/Fan module or HVAC system required
Optical system – Excitation & Detection	In-capillary dual-side laser excitation & CCD camera detection	Scanning confocal optics for in-capillary laser excitation & detection by PMTs
Instrument accessories	- UPS for backup during power outages - PC \ Monitor \ Barcode Scanner	- UPS for backup during power outages - PC \ Monitor \ Barcode Scanner - High & Low pressure N2 systems
Reagents/Materials Handling	- Automated sieving matrix delivery pump & on-board reagents for 2-day unattended operation. - Integrated auto-sampler & sample plate stacker (16 plate capacity)	- Technician loaded sieving matrix & reagents for each run performed. - Individually loaded sample plates by Technician
Sequencing Chemistry	BigDye v3	DYEnamic ET dye terminators
Instrument Software	Unified Data Collection v3.0	Instrument Control Manager v4.2
Basecalling Software	KB Basecaller	Sequence Analyzer v4.0 with Cimarron 3.12 basecaller

JGI Throughput and Read Length Comparison of 36 MB4500s vs. 71 ABI3730s



Conclusion

In 2006 the JGI Production Line produced 49 million sequencing reads and 33 billion base pairs of sequence. The Production Line maintains this level of production through the integration of several groups into the overall operation. The groups that ensure that the Production Line performs at a throughput of up to 35 billion bases per year are Reagent Quality Control, Data Quality Assurance, and Instrumentation. These groups ensure that the Production Line and equipment are operating within acceptable specifications. Using continuous improvement principles, the Production Department continuously re-evaluates its systems and processes and strives to remove unnecessary steps in the workflow, reduce reagent usage, improve ergonomics, and reduce cost overall, while maintaining the production of high quality DNA sequence.

UCRL-POST-229398

LBNL-62525 Poster

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.